

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: NOVO NORDISK A/S, N N
- (ii) TITLE OF INVENTION: A Cellulase Preparation
- 10 (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: NOVO NORDISK A/S, Patent Department
- 15 (B) STREET: Novo Alle
- (C) CITY: Bagsvaerd
- (E) COUNTRY: DENMARK
- (F) ZIP: DK-2880
- (v) COMPUTER READABLE FORM:
- 20 (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- 25 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:
- 30 (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Thalsee-Madsen, Birgit
- (ix) TELECOMMUNICATION INFORMATION:
- 35 (A) TELEPHONE: +45 4444 8888
- (B) TELEFAX: +45 4449 3256
- (C) TELEX: 37304

(2) INFORMATION FOR SEQ ID NO:1:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1060 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 45 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- 50 (vi) ORIGINAL SOURCE:
- (A) ORGANISM: Humicola insolens
- (B) STRAIN: DSM 1800

(ix) FEATURE:

(A) NAME/KEY: mat_peptide

(B) LOCATION: 73..927

5 (ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 10..72

10 (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 10..927

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

15 GGATCCAAG ATG OGT TOC TOC CCC CTC CTC CCG TOC GCC GGT GIG GCC 48
Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala
-21 -20 -15 -10

20 GCC CTG CCG GTG TTG GCC CTT GCC GCT GAT GGC AGG TOC ACC CGC TAC 96
Ala Leu Pro Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr
-5 1 5

TGG GAC TGC TGC AAG CCT TOG TGC GGC TGG GGC AAG AAG GCT CCC GTG 144
25 Trp Asp Cys Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val
10 15 20

AAC CAG CCT GTC TTT TOC TGC AAC GCC AAC TTC CAG CGT ATC ACG GAC 192
30 25 Asn Gln Pro Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp
30 35 40

TTC GAC GCC AAG TOC GGC TGC GAG CCG GGC GGT GTC GCC TAC TCG TGC 240
Phe Asp Ala Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys
45 50 55

35 GCC GAC CAG ACC CCA TGG GCT GTG AAC GAC GAC TTC GCG CTC GGT TTT 288
Ala Asp Gln Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe
60 65 70

40 GCT GCC ACC TCT ATT GCC GGC AGC AAT GAG GCG GGC TGG TGC TGC GCC 336
Ala Ala Thr Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala
75 80 85

TGC TAC GAG CTC ACC TTC ACA TOC GGT CCT GTT GCT GGC AAG AAG ATG 384
45 Cys Tyr Glu Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met
90 95 100

GTC GTC CAG TOC ACC AGC ACT GGC GGT GAT CTT GGC AGC AAC CAC TTC 432
50 105 Val Val Gln Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe
110 115 120

GAT CTC AAC ATC CCC GGC GGC GGC GTC GGC ATC TTC GAC GGA TGC ACT 480
Asp Leu Asn Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr
125 130 135

[illegible]

45

[illegible]

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 305 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala Ala Leu Pro
-21 -20 -15 -10

15 Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
-5 1 5 10

Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
15 20 25

20 Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
30 35 40

Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
25 45 50 55

Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
60 65 70 75

30 Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
80 85 90

Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
95 100 105

35 Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
110 115 120

Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
40 125 130 135

Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
140 145 150 155

45 Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
160 165 170

Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
175 180 185

50 Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
190 195 200

Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser
55 205 210 215

00735767-14300

	AGC TGG TGC TGT GCT TGC TAT GCT TTG ACC TTC ACC ACT GGC CCC GTC	450
	Ser Trp Cys Cys Ala Cys Tyr Ala Leu Thr Phe Thr Thr Gly Pro Val	
	105 110 115	
5	AAG GGC AAG AAG ATG ATC GTC CAG TCC ACC AAC ACT GGA GGT GAT CTC	498
	Lys Gly Lys Lys Met Ile Val Gln Ser Thr Asn Thr Gly Gly Asp Leu	
	120 125 130	
10	GGC GAC AAC CAC TTC GAT CTC ATG ATG CCC GGC GGT GGT GTC GGT ATC	546
	Gly Asp Asn His Phe Asp Leu Met Met Pro Gly Gly Gly Val Gly Ile	
	135 140 145 150	
15	TTC GAC GGC TGC ACC TCT GAG TTC GGC AAG GCT CTC GGC GGT GCC CAG	594
	Phe Asp Gly Cys Thr Ser Glu Phe Gly Lys Ala Leu Gly Gly Ala Gln	
	155 160 165	
20	TAC GGC GGT ATC TCC TCC CGA ACC GAA TGT GAT AGC TAC CCC GAG CTT	642
	Tyr Gly Gly Ile Ser Ser Arg Ser Glu Cys Asp Ser Tyr Pro Glu Leu	
	170 175 180	
25	CTC AAG GAC GGT TGC CAC TGG CGA TTC GAC TGG TTC GAG AAC GCC GAC	690
	Leu Lys Asp Gly Cys His Trp Arg Phe Asp Trp Phe Glu Asn Ala Asp	
	185 190 195	
30	AAC CCT GAC TTC ACC TTT GAG CAG GTT CAG TGC CCC AAG GCT CTC CTC	738
	Asn Pro Asp Phe Thr Phe Glu Gln Val Gln Cys Pro Lys Ala Leu Leu	
	200 205 210	
35	GAC ATC AGT GGA TGC AAG CGT GAT GAC GAC TCC AGC TTC OCT GCC TTC	786
	Asp Ile Ser Gly Cys Lys Arg Asp Asp Ser Ser Phe Pro Ala Phe	
	215 220 225 230	
40	AAG GTT GAT ACC TCG GCC AGC AAG CCC CAG CCC TCC AGC TCC GCT AAG	834
	Lys Val Asp Thr Ser Ala Ser Lys Pro Gln Pro Ser Ser Ser Ala Lys	
	235 240 245	
45	AAG ACC ACC TCC GCT GCT GCT GCC GCT CAG CCC CAG AAG ACC AAG GAT	882
	Lys Thr Thr Ser Ala Ala Ala Ala Ala Gln Pro Gln Lys Thr Lys Asp	
	250 255 260	
50	TCC GCT OCT GTT GTC CAG AAG TCC TCC ACC AAG OCT GCC GCT CAG CCC	930
	Ser Ala Pro Val Val Gln Lys Ser Ser Thr Lys Pro Ala Ala Gln Pro	
	265 270 275	
55	GAG OCT ACT AAG CCC GCC GAC AAG CCC CAG ACC GAC AAG OCT GTC GCC	978
	Glu Pro Thr Lys Pro Ala Asp Lys Pro Gln Thr Asp Lys Pro Val Ala	
	280 285 290	
60	ACC AAG OCT GCT GCT ACC AAG CCC GTC CAA OCT GTC AAC AAG CCC AAG	1026
	Thr Lys Pro Ala Ala Thr Lys Pro Val Gln Pro Val Asn Lys Pro Lys	
	295 300 305 310	
65	ACA ACC CAG AAG GTC CGT GGA ACC AAA ACC CGA GGA AGC TGC CCG GCC	1074
	Thr Thr Gln Lys Val Arg Gly Thr Lys Thr Arg Gly Ser Cys Pro Ala	
	315 320 325	

G0915707-14300

AAG ACT GAC GCT ACC GGC AAG GGC TCC GGT GTC CCT GCT TAT TAC CAG 1122
 Lys Thr Asp Ala Thr Ala Lys Ala Ser Val Val Pro Ala Tyr Tyr Gln
 330 335 340

5 TGT GGT GGT TCC AAG TCC GCT TAT CCC AAC GGC AAC CTC GCT TGC GCT 1170
 Cys Gly Gly Ser Lys Ser Ala Tyr Pro Asn Gly Asn Leu Ala Cys Ala
 345 350 355

ACT GGA AGC AAG TGT GTC AAG CAG AAC GAG TAC TAC TCC CAG TGT GTC 1218
 10 Thr Gly Ser Lys Cys Val Lys Gln Asn Glu Tyr Tyr Ser Gln Cys Val
 360 365 370

CCC AAC TAAATGGTAG ATCCATCGGT TTGGGAAGAG ACTATGGGTC TCAGAAGGGA 1274
 Pro Asn
 15 375

TCCTCTCATG AGCAGGCTTG TCATGTGATA GCATGGCATC CTTGACCAAG TGTTCGACCC 1334

TGTGTGTACA TAGTATATCT TCATGTGATA TATTAGACA CATAGATAGC CTCCTGTGAG 1394
 20

CGACAACTGG CTACAAAAGA CTGGCAGGC TTGTTCATA TTGACACAGT TTCTCCATA 1454

AAAAAAAAAA AAAAAAAAAA 1473

00735797-101300

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 376 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Ser Tyr Thr Leu Leu Ala Leu Ala Gly Pro Leu Ala Val Ser
1 5 10 15
Ala Ala Ser Gly Ser Gly His Ser Thr Arg Tyr Trp Asp Cys Cys Lys
20 25 30
Pro Ser Cys Ser Trp Ser Gly Lys Ala Ala Val Asn Ala Pro Ala Leu
20 35 40 45
Thr Cys Asp Lys Asn Asp Asn Pro Ile Ser Asn Thr Asn Ala Val Asn
50 55 60
25 Gly Cys Glu Gly Gly Gly Ser Ala Tyr Ala Cys Thr Asn Tyr Ser Pro
65 70 75 80
Trp Ala Val Asn Asp Glu Leu Ala Tyr Gly Phe Ala Ala Thr Lys Ile
85 90 95
30 Ser Gly Gly Ser Glu Ala Ser Trp Cys Cys Ala Cys Tyr Ala Leu Thr
100 105 110
Phe Thr Thr Gly Pro Val Lys Gly Lys Lys Met Ile Val Gln Ser Thr
35 115 120 125
Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Met Met Pro
130 135 140
40 Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Ser Glu Phe Gly Lys
145 150 155 160
Ala Leu Gly Gly Ala Gln Tyr Gly Gly Ile Ser Ser Arg Ser Glu Cys
165 170 175
45 Asp Ser Tyr Pro Glu Leu Leu Lys Asp Gly Cys His Trp Arg Phe Asp
180 185 190
Trp Phe Glu Asn Ala Asp Asn Pro Asp Phe Thr Phe Glu Gln Val Gln
50 195 200 205
Cys Pro Lys Ala Leu Leu Asp Ile Ser Gly Cys Lys Arg Asp Asp Asp
210 215 220
55 Ser Ser Phe Pro Ala Phe Lys Val Asp Thr Ser Ala Ser Lys Pro Gln
225 230 235 240

007357-130

(1) GENERAL INFORMATION:

- (i) APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Moller
Schulein, Martin
Patkar, Shankant A.
Hagen, Fred
- (ii) TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme
- (iii) NUMBER OF SEQUENCES: 33
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Novo Nordisk of North America, Inc.
 - (B) STREET: 405 Lexington Avenue, 64th Floor
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: United States of America
 - (F) ZIP: 10174-6401
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/389,423
 - (B) FILING DATE: 14-FEB-1995
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Lambiris, Elias J.
 - (B) REGISTRATION NUMBER: 33,728
 - (C) REFERENCE/DOCKET NUMBER: 3469.214-US
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 212-867-0123
 - (B) TELEFAX: 212-878-9655

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1060 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

62
48

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Humicola insolens*
 (B) STRAIN: DSM 1800

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 73..924

(ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 10..72

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 10..924

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGATCCAAG ATG CGT TCC TCC CCC CTC CTC CCG TCC GCC GTT GTG GCC	48
Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala	
-21 -20 -15 -10	
GCC CTG CCG GTG TTG GCC CTT GCC GCT GAT GGC AGG TCC ACC CGC TAC	96
Ala Leu Pro Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr	
-5 1 5	
TGG GAC TGC TGC AAG CCT TCG TGC GGC TGG GCC AAG AAG GCT CCC GTG	144
Trp Asp Cys Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val	
10 15 20	
AAC CAG CCT GTC TTT TCC TGC AAC GCC AAC TTC CAG CGT ATC ACG GAC	192
Asn Gln Pro Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp	
25 30 35 40	
TTC GAC GCC AAG TCC GGC TGC GAG CCG GGC GGT GTC GCC TAC TCG TGC	240
Phe Asp Ala Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys	
45 50 55	
GCC GAC CAG ACC CCA TGG GCT GTG AAC GAC GAC TTC GCG CTC GGT TTT	288
Ala Asp Gln Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe	
60 65 70	
GCT GCC ACC TCT ATT GCC GGC AGC AAT GAG GCG GGC TGG TGC TGC GCC	336
Ala Ala Thr Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala	
75 80 85	
TGC TAC GAG CTC ACC TTC ACA TCC GGT CCT GTT GCT GGC AAG AAG ATG	384
Cys Tyr Glu Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met	
90 95 100	
GTC GTC CAG TCC ACC AGC ACT GGC GGT GAT CTT GGC AGC AAC CAC TTC	432
Val Val Gln Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe	
105 110 115 120	
GAT CTC AAC ATC CCC GGC GGC GGC GTC GGC ATC TTC GAC GGA TGC ACT	480

Asp	Leu	Asn	Ile	Pro	Gly	Gly	Gly	Val	Gly	Ile	Phe	Asp	Gly	Cys	Thr	
				125					130					135		
CCC	CAG	TTC	GGC	GGT	CTG	CCC	GGC	CAG	CGC	TAC	GGC	GGC	ATC	TCG	TCC	528
Pro	Gln	Phe	Gly	Gly	Leu	Pro	Gly	Gln	Arg	Tyr	Gly	Gly	Ile	Ser	Ser	
			140					145					150			
CGC	AAC	GAG	TGC	GAT	CGG	TTC	CCC	GAC	GCC	CTC	AAG	CCC	GGC	TGC	TAC	576
Arg	Asn	Glu	Cys	Asp	Arg	Phe	Pro	Asp	Ala	Leu	Lys	Pro	Gly	Cys	Tyr	
		155					160					165				
TGG	CGC	TTC	GAC	TGG	TTC	AAG	AAC	GCC	GAC	AAT	CCG	AGC	TTC	AGC	TTC	624
Trp	Arg	Phe	Asp	Trp	Phe	Lys	Asn	Ala	Asp	Asn	Pro	Ser	Phe	Ser	Phe	
		170				175					180					
CGT	CAG	GTC	CAG	TGC	CCA	GCC	GAG	CTC	GTC	GCT	CGC	ACC	GGA	TGC	CGC	672
Arg	Gln	Val	Gln	Cys	Pro	Ala	Glu	Leu	Val	Ala	Arg	Thr	Gly	Cys	Arg	
185					190					195				200		
CGC	AAC	GAC	GAC	GGC	AAC	TTC	CCT	GCC	GTC	CAG	ATC	CCC	TCC	AGC	AGC	720
Arg	Asn	Asp	Asp	Gly	Asn	Phe	Pro	Ala	Val	Gln	Ile	Pro	Ser	Ser	Ser	
				205					210					215		
ACC	AGC	TCT	CCG	GTC	AAC	CAG	CCT	ACC	AGC	ACC	AGC	ACC	ACG	TCC	ACC	768
Thr	Ser	Ser	Pro	Val	Asn	Gln	Pro	Thr	Ser	Thr	Ser	Thr	Thr	Ser	Thr	
			220				225					230				
TCC	ACC	ACC	TCG	AGC	CCG	CCA	GTC	CAG	CCT	ACG	ACT	CCC	AGC	GGC	TGC	816
Ser	Thr	Thr	Ser	Ser	Pro	Pro	Val	Gln	Pro	Thr	Thr	Pro	Ser	Gly	Cys	
			235				240					245				
ACT	GCT	GAG	AGG	TGG	GCT	CAG	TGC	GGC	GGC	AAT	GGC	TGG	AGC	GGC	TGC	864
Thr	Ala	Glu	Arg	Trp	Ala	Gln	Cys	Gly	Gly	Asn	Gly	Trp	Ser	Gly	Cys	
	250					255				260						
ACC	ACC	TGC	GTC	GCT	GGC	AGC	ACT	TGC	ACG	AAG	ATT	AAT	GAC	TGG	TAC	912
Thr	Thr	Cys	Val	Ala	Gly	Ser	Thr	Cys	Thr	Lys	Ile	Asn	Asp	Trp	Tyr	
265					270					275				280		
CAT	CAG	TGC	CTG	TAGACGCAGG	GCAGCTTGAG	GGCCTTACTG	GTGGCCGCAA									964
His	Gln	Cys	Leu													
CGAAATGACA	CTCCCAATCA	CTGTATTAGT	TCTTGACAT	AATTCGTCA	TCCCTCCAGG											1024
GATTGTCACA	TAAATGCAAT	GAGGAACAAT	GAGTAC													1060

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 305 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys
270 275 280

Leu

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Fusarium oxysporum*
(B) STRAIN: DSM 2672

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 97..1224

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATTCGCGG	CCGCTCATTC	ACTTCATTCA	TTCTTTAGAA	TTACATACAC	TCTCTTTCAA		60									
AACAGTCACT	CTTTAAACAA	AACAACTTTT	GCAACA	ATG	CGA	TCT	TAC	ACT	CTT		114					
				Met	Arg	Ser	Tyr	Thr	Leu							
				1					5							
CTC	GCC	CTG	GCC	GGC	CCT	CTC	GCC	GTG	AGT	GCT	GCT	TCT	GGA	AGC	GGT	162
Leu	Ala	Leu	Ala	Gly	Pro	Leu	Ala	Val	Ser	Ala	Ala	Ser	Gly	Ser	Gly	
			10					15					20			
CAC	TCT	ACT	CGA	TAC	TGG	GAT	TGC	TGC	AAG	CCT	TCT	TGC	TCT	TGG	AGC	210
His	Ser	Thr	Arg	Tyr	Trp	Asp	Cys	Cys	Lys	Pro	Ser	Cys	Ser	Trp	Ser	
		25					30					35				
GGA	AAG	GCT	GCT	GTC	AAC	GCC	CCT	GCT	TTA	ACT	TGT	GAT	AAG	AAC	GAC	258
Gly	Lys	Ala	Ala	Val	Asn	Ala	Pro	Ala	Leu	Thr	Cys	Asp	Lys	Asn	Asp	
	40					45					50					
AAC	CCC	ATT	TCC	AAC	ACC	AAT	GCT	GTC	AAC	GGT	TGT	GAG	GGT	GGT	GGT	306
Asn	Pro	Ile	Ser	Asn	Thr	Asn	Ala	Val	Asn	Gly	Cys	Glu	Gly	Gly	Gly	
	55				60					65					70	
TCT	GCT	TAT	GCT	TGC	ACC	AAC	TAC	TCT	CCC	TGG	GCT	GTC	AAC	GAT	GAG	354
Ser	Ala	Tyr	Ala	Cys	Thr	Asn	Tyr	Ser	Pro	Trp	Ala	Val	Asn	Asp	Glu	
				75					80					85		

67
33

ACA ACC CAG AAG GTC CGT GGA ACC AAA ACC CGA GGA AGC TGC CCG GCC	1074
Thr Thr Gln Lys Val Arg Gly Thr Lys Thr Arg Gly Ser Cys Pro Ala	
315 320 325	
AAG ACT GAC GCT ACC GCC AAG GCC TCC GTT GTC CCT GCT TAT TAC CAG	1122
Lys Thr Asp Ala Thr Ala Lys Ala Ser Val Val Pro Ala Tyr Tyr Gln	
330 335 340	
TGT GGT GGT TCC AAG TCC GCT TAT CCC AAC GGC AAC CTC GCT TGC GCT	1170
Cys Gly Gly Ser Lys Ser Ala Tyr Pro Asn Gly Asn Leu Ala Cys Ala	
345 350 355	
ACT GGA AGC AAG TGT GTC AAG CAG AAC GAG TAC TAC TCC CAG TGT GTC	1218
Thr Gly Ser Lys Cys Val Lys Gln Asn Glu Tyr Tyr Ser Gln Cys Val	
360 365 370	
CCC AAC TAAATGGTAG ATCCATCGGT TGTGGAAGAG ACTATGCGTC TCAGAAGGGA	1274
Pro Asn	
375	
TCCTCTCATG AGCAGGCTTG TCATTGTATA GCATGGCATC CTGGACCAAG TGTTGACCC	1334
TTGTTGTACA TAGTATATCT TCATTGTATA TATTTAGACA CATAGATAGC CTCTTGTGAC	1394
CGACAACCTGG CTACAAAAGA CTTGGCAGGC TTGTTCAATA TTGACACAGT TTCCTCCATA	1454
AAAAAAAAAA AAAAAAAAAA	1473

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 376 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Ser Tyr Thr Leu Leu Ala Leu Ala Gly Pro Leu Ala Val Ser	
1 5 10 15	
Ala Ala Ser Gly Ser Gly His Ser Thr Arg Tyr Trp Asp Cys Cys Lys	
20 25 30	
Pro Ser Cys Ser Trp Ser Gly Lys Ala Ala Val Asn Ala Pro Ala Leu	
35 40 45	
Thr Cys Asp Lys Asn Asp Asn Pro Ile Ser Asn Thr Asn Ala Val Asn	
50 55 60	
Gly Cys Glu Gly Gly Gly Ser Ala Tyr Ala Cys Thr Asn Tyr Ser Pro	
65 70 75 80	
Trp Ala Val Asn Asp Glu Leu Ala Tyr Gly Phe Ala Ala Thr Lys Ile	

[illegible]

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEO ID NO:5:

AGCTGCGGCC GCAGGCCGCG GAGGCCA

27

(2) INFORMATION FOR SEO ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCTTGGCCT CCGCGGCCTG CGGCCGC

27

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AATTCGCGGC CGCGGCCATG GAGGCC

26

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

20

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

[illegible]

71
57

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCGCCAGTAG CAGCCGGGCT TGAGGG

26

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ACGTCTCAAC TCGGATCCAA GATGCGTT

28

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCAACTCTG ATCAAGATGC GTTCC

25

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TGTCGACCAG TAAGGCCCTC AAGCTG

26

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GACAGAGCAC AGAATTCACT AGTGAGCTCT

30

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEO ID NO:17:

TGGGAYTGYT GYAARCC

17

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEO ID NO:18:

AGGGAGACCG GAATTCTGGG AYTGYTGYAA RCC

33

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

```
(C) STRANDEDNESS: single
```

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCNGGNGGNG GNGTNGG

17

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGGGAGACCG GAATTCCTCCNG GNGGNGGNGT NGG

33

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACNAYCATNK TYTTNCC

17

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GACAGAGCAC AGAATTCACN AYCATTNKTYT TNCC

34

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

$$\begin{array}{r} 74 \\ \underline{60} \end{array}$$
[illegible]

- (ii) MOLECULE TYPE: cDNA

GACAGAGCAC AGAATTCNGG RTTRTCNGCN KYITYRAACC A

41

(ii) MOLECULE TYPE: cDNA

GGGGTAGCTA TCACATTCGC TTCGGGAGGA GATACCGCCG TA

42

(ii) MOLECULE TYPE: cDNA

CTTCTTGCTC TTGGAGCGGA AAGGCTGCTG TCAACGCCCC TG

42

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGTACGCATG TAACATTA

18

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTGCACAATA TTTCAAGC

18

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGGGTAGCTA TCACATTCGC TTCGGGAGGA GATACCGCCG TA

42

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTTCTTGCTC TTGGAGCGGA AAGGCTGCTG TCAACGCCCC TG

42

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs

76
62

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AGCTTCTCAA GGACGGTT

18

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AACAAGGGTC GAACACTT

18

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CCAGAAGACC AAGGATT

17

0075707-4075707